1.0

4.5
5.0
5.6

2.8 2.5

3.2

2.2

3.6

4.0

2.0

1.1

1.8

1.25 1.4 1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

University of Missouri-Columbia

A Bayesian Nonparametric Approach to Reliability

by

R. L. Dykstra and Purushottam Laud

Technical Report No. 84
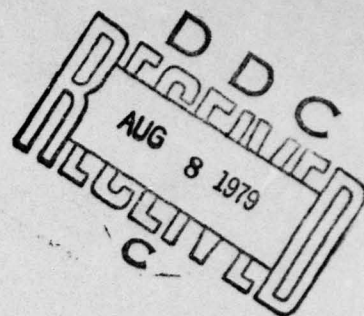Department of Statistics   July 1979

Mathematical Sciences

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 84	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* A Bayesian Nonparametric Approach to Reliability.		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard L. Dykstra and Purushottam Laud		8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0655
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of Missouri Columbia, Missouri 65211		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, Virginia		12. REPORT DATE July 10, 1979
		13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

hazard rates; increasing hazard rates; decreasing hazard rates; Bayes estimates; extended gamma process; posterior process; prior process; nonparametric

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

It is suggested that problems in a reliability context may be handled by a Bayesian non-parametric approach. A stochastic process is defined whose sample paths may be assumed to be either increasing hazard rates or decreasing hazard rates by → next page

79 08 7 011

properly choosing the parameter functions of the process. The
posterior distribution of the hazard rates are derived for both
exact and censored data. Bayes estimates of hazard rates,
c.d.f.'s, densities, and means, are found under squared error type
loss functions. Some simulation is done and estimates graphed to
better understand the estimators. Finally, estimates of the
c.d.f. from some data in a paper by Kaplan and Meier are
constructed.

Accession For

NTIS GRA&I

DDC TAB

Unannounced

Justification_____

By_____

Distribution/

Availability Codes

Dist | Avail and/or
special

A

A Bayesian Nonparametric
Approach to Reliability

by

R. L. Dykstra and Purushottam Laud

Technical Repart No. 84
July 1979

Department of Statistics
University of Missouri
Columbia, Missouri 65211

SUMMARY

It is suggested that problems in a reliability context may be handled by a Bayesian non-parametric approach. A stochastic process is defined whose sample paths may be assumed to be either increasing hazard rates or decreasing hazard rates by properly choosing the parameter functions of the process. The posterior distribution of the hazard rates are derived for both exact and censored data. Bayes estimates of hazard rates, c.d.f.'s, densities, and means, are found under squared error type loss functions. Some simulation is done and estimates graphed to better understand the estimators. Finally, estimates of the c.d.f from some data in a paper by Kaplan and Meier are constructed.

1. INTRODUCTION

Recently, there has been a good deal of interest in nonparametric Bayesian approaches to statistical inference. In this approach a stochastic process is defined whose sample paths index a large family of distributions. Then conditional on a realization of the process, i.i.d. observations are taken from the indexed distribution, and inferences are made from the posterior distribution of the process. In this manner the prior probability can be spread over a very large number of distributions. It is also possible to avoid explicitly specifying the functional form of the likelihood.

The most common approach has been extensively discussed by Ferguson (1973), and consists of using a 'Dirichlet Process' prior. That is, a continuous time parameter stochastic process whose finite dimensional increments have a Dirichlet distribution is defined. One can then assume that the sample paths of this process are cumulative distribtion functions. Ferguson shows that the posterior distribution of the process, given the complete observations, is also distributed as a Dirichlet stochastic process, and uses this posterior distribution for making his statistical inferences.

Antoniak (1974) considers mixtures of Dirichlet distributions. Doksum (1974) addresses his attention to prior stochastic processes that are 'tailfree', and/or 'neutral'. His posterior distributions, however, are obtained in terms of expectations over the entire probability space. Susarla and Van Ryzin (1976) were able to obtain the posterior mean of censored data using a Dirichlet prior. Recently, Ferguson and Phadia (1976) were able to generalize these censored data results to more general "neutral to the right" processes.

This type of approach seems to have merit concerning statistical inference in a reliability context. What is proposed, since the concept of hazard rate plays such a key role in statistical reliability, is to place the prior probability over the collection of hazard rates. This is done by defining an appropriate stochastic process whose sample paths are hazard rates. With this prior we derive the posterior distribution of the hazard rates for both right censored and exact observations. This approach has the advantage of placing the prior probability strictly on absolutely continuous distributions rather than on discrete distributions as is the case with the Dirichlet process prior. Moreover, Bayes estimators of the entire distribution under natural loss functions are absolutely continuous. Finally, since our prior random c.d.f.'s are not neutral to the right, the work of Doksum (1974) and Ferguson and Phadia (1976) does not apply.

2. THE EXTENDED GAMMA PROCESS

We shall assume throughout that our distributions have positive probability only on the nonnegative half of the real line, although one could adapt to distributions over the whole real line. The hazard function $H(x)$ of a distribution is defined to be

$$H(x) = - \ln(1 - F(x))$$

where $F(x)$ is the left continuous c.d.f. of the distribution as in Loeve (1963). (It is also possible to work with right continuous c.d.f.'s, but left continuous c.d.f.'s are computationally more convenient for us.) We shall refer to $\overline{F}(x) = 1 - F(x)$ as

the survival function of the distribution. Note that from some point on, $H(x)$ may equal plus infinity. If, for all x, one can express

$$H(x) = \int_{[0,x)} r(t) \; dt \; ,$$

then $r(x)$ is called the hazard rate of the distribution. Thus, $r(x)$ is related to the density $f(x)$ by the relationship

$$r(x) = \frac{f(x)}{\overline{F}(x)} \; ,$$

and has the interpretation that $r(x)\Delta$ is approximately equal to the probability of failure in the next Δ increment of time given that a lifetime has survived until time x.

We denote by $G(\alpha,\beta)$ the gamma distribution with shape parameter $\alpha \geq 0$, and scale parameter $\beta > 0$. For $\alpha > 0$, this distribution has for its density with respect to Lebesgue measure,

$$g(x|\alpha,\beta) = x^{\alpha-1} \exp(-x/\beta) I_{(0,\infty)}(x)/\Gamma(\alpha)\beta^{\alpha} \; ,$$

with the distribution assumed to be degenerate at 0 if $\alpha = 0$.

Let $\alpha(t), t \geq 0$, be a nondecreasing left-continuous real-valued function such that $\alpha(0) = 0$, and let $\beta(t), t \geq 0$, be a positive right-continuous real-valued function, bounded away from 0 and ∞ with left hand limits existing.

$Z(t)$, $t \geq 0$, defined on an appropriate probability space $(\Omega, \mathscr{F}, p)$ denotes a gamma process with independent increments corresponding to $\alpha(t)$. That is, $Z(0) \equiv 0$, $Z(t)$ has independent increments and for $t > s$, $Z(t) - Z(s)$ is $G(\alpha(t) - \alpha(s), 1)$.

It has been shown (see Ferguson (1973) that such a process exists and that its distribution is uniquely determined. We assume WLOG that this process has nondecreasing left continuous sample paths.

We now define a new stochastic process by

$$(2.1) \qquad r(t) = \int_{[0,t)} \beta(s) dZ(s) ,$$

with the interpretation that for almost every ω, $Z(t,\omega)$ is a nondecreasing left continuous function in t bounded on every finite interval, and $r(t)$ is the Lebesque Stieltjes integral, with respect to that function, of $\beta(s)$ over the interval $[0,t)$.

We say a process defined in this manner has an extended gamma distribution, and we denote such a process by

$$r(t) \quad \text{is} \quad \Gamma(\alpha(\cdot), \beta(\cdot)) .$$

The finite dimensional c.d.f.'s (or densities) of $r(t)$ appear to be rather intractable, although the distribution of the extended gamma process is "nice" in many ways.

THEOREM 2.1   If $r(t)$ is distributed as $\Gamma(\alpha(\cdot), \beta(\cdot))$, then $r(t)$ has independent increments and for fixed t

(2.2)   the characteristic function of $r(t)$ in some

neighborhood of 0 is given by

$$\psi_{r(t)}(\theta) = \exp\left[-\int_{[0,t)} \ln(1 - i\beta(s)\theta)d\alpha(s)\right]$$

(2.3) $\displaystyle Er(t) = \int_{[0,t)} \beta(s)d\alpha(s)$, and

(2.4) $\displaystyle Var\ r(t) = \int_{[0,t)} \beta^2(s)d\alpha(s)$.

<u>PROOF</u>:   Let $0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{k(n)}^{(n)}$ be a sequence of parti-

tions whose norm goes to 0 and $t_{k(n)}^{(n)} \to \infty$ as $n \to \infty$. Define

(2.5) $\displaystyle r_n(t) = \sum_{\{i>0;\, t_i^{(n)} < t\}} \beta(t_i^{(n)})\left[Z(t_i^{(n)}) - Z(t_{i-1}^{(n)})\right]$

where if the index set is empty, we assume $r(t) \equiv 0$. Then
$r_n(t) \xrightarrow{a.s.} r(t)$ so that $r_n(t) \xrightarrow{\ell} r(t)$ and $\psi_{r_n(t)}(\theta) \to \psi_{r(t)}(\theta)$.

Also, $\displaystyle \psi_{r_n(t)}(\theta) = \prod_{\{j;\, t_j^{(n)} < t\}} \psi_{Z(t_j^{(n)}) - Z(t_{j-1}^{(n)})}(\beta(t_j^{(n)})\theta)$

$\displaystyle = \prod_{\{j;\, t_j^{(n)} < t\}} (1 - i\beta(t_j^{(n)})\theta)^{-(\alpha(t_j^{(n)}) - \alpha(t_{j-1}^{(n)}))}$

$\displaystyle = \exp\left[-\sum_{\{j;\, t_j^{(n)} < t\}} (\alpha(t_j^{(n)}) - \alpha(t_{j-1}^{(n)}))\ln(1 - i\beta(t_j^{(n)})\theta)\right]$

$\displaystyle \to \exp\left[-\int_{[0,t)} \ln(1 - i\beta(s)\theta)d\alpha(s)\right]$ for θ sufficiently close

to 0.

One can show (2.3) and (2.4) either by taking derivatives of the characteristic function or by verifying that $Er_n(t)^k \to Er(t)^k$, $k > 0$. The independent increments follow easily by letting the increment endpoints be contained in the partition points.

Since the original gamma process $Z(t)$ is a pure jump process, the extended gamma process will also be a pure jump process.

3. RANDOM HAZARD RATES

Provided $\alpha(t)$ is not identically zero, we may assume that the sample paths of an extended gamma process $r(t)$ are well defined nondecreasing hazard rates corresponding to absolutely continuous distributions. Thus the conditional distribution of the observations $X_1, \ldots, X_n$ given $r(t)$ will be defined by

$$(3.1) \quad P(X_1 \geq x_1, \ldots, X_n \geq x_n | r(t)) = \prod_{i=1}^{n} \exp\left[-\int_{[0,x_i)} r(t)dt\right] \text{ a.s.}$$

Of course (3.1) and the distribution of $r(t)$ will determine the joint distribution of $X_1, \ldots, X_n$, $r(t)$ and will be used to derive the marginal distribution of $X_1, \ldots, X_n$ and the posterior distribution of $r(t)$ given the observed values of $X_1, \ldots, X_n$. Since the sample paths of the $r(t)$ process are nondecreasing functions a.s., we are placing our prior probability entirely within the class of distributions with nondecreasing hazard rates. Later, we will show how the prior can be placed over distributions with nonincreasing hazard rates.

In assigning a prior probability measure by this method, one

needs to input the functions $\alpha(t)$ and $\beta(t)$. One approach consists of defining nondecreasing mean and variance functions $\mu(t)$ and $\sigma^2(t)$. It would seem reasonable to assign as $\mu(t)$ the best "guess" of the hazard rate and use $\sigma^2(t)$ to measure the amount of uncertainty or variation in the hazard rate at the point t. Thus a band $\mu(t) \pm 2\sigma(t)$ should cover most of the "feeling" for the location of the hazard rate. Assuming $\mu(t)$, $\sigma^2(t)$ and $\sigma(t)$ are all differentiable, one can use (2.3) and (2.4) to set

$$\mu(t) = \int_{[0,t)} \beta(s)\alpha'(s)ds, \quad \text{and}$$

$$\sigma^2(t) = \int_{[0,t)} \beta^2(s)\alpha'(s)ds .$$

Solving for $\alpha(t)$ and $\beta(t)$ yields

$$(3.2) \qquad \beta(t) = \frac{d\sigma^2(t)}{dt} \Big/ \frac{d\mu(t)}{dt} , \quad \text{and}$$

$$(3.3) \qquad \frac{d\alpha(t)}{dt} = \frac{d\mu(t)^2}{dt} \Big/ \frac{d\sigma^2(t)}{dt} ,$$

which then determines the prior distribution. The form of the posterior distribution gives information on the effect of the prior and may help in choosing $\alpha(\cdot)$ and $\beta(\cdot)$.

The marginal distribution of an observation X can be found from (3.1) with the use of a limiting argument. The proof of Theorem 3.1 is given in section 7.

THEOREM 3.1   If the prior over hazard rates is $\Gamma(\alpha(\cdot), \beta(\cdot))$ then the marginal survival function of an observation X is given by

$$(3.4) \quad \overline{F}(t) = P(X \geq t) = \exp\left[- \int_{[0,t)} \ln(1+\beta(s)(t-s)) d\alpha(s)\right].$$

The marginal survival function of the observations $X_1, \ldots, X_n$ can be found by methods similar to Theorem 3.1 and is given in the following corollary.

COROLLARY 3.1   If the prior over the hazard rates is $\Gamma(\alpha(\cdot), \beta(\cdot))$, then the joint marginal survival function of n observations $X_1, \ldots, X_n$ is

$$(3.5) \quad F(t_1, \ldots, t_n) = P(X_1 \geq t_1, \ldots, X_n \geq t_n) = \exp\left[- \int_{[0,\infty)} \ln\left(1+\beta(s) \sum_{i=1}^{n} (s-t_i)^+\right) d\alpha(s)\right]$$

where $a^+ = \sup\{a, 0\}$.

Thus the marginal survival function of $Y = \min(X_1, \ldots, X_n)$ is of the same form as the survival function of just X_1 providing $\beta(s)$ is replaced by $n\beta(s)$.

The key problem in any Bayesian setting is to derive the posterior distribution. Moreover it is important to handle censored observations since reliability data are often of this type. If an extended gamma prior is used, the posterior distribution for right censored observations is also an extended gamma process. The proof is given in section 7.

THEOREM 3.2   If the prior over the hazard rates is $\Gamma(\alpha(\cdot), \beta(\cdot))$ then the posterior over the hazard rates, given m censored observations of the form $X_1 \geq x_1, X_2 \geq x_2, \ldots, X_m \geq x_m$ is $\Gamma(\alpha(\cdot), \hat{\beta}(\cdot))$ where

$$(3.6) \qquad \hat{\beta}(t) = \frac{\beta(t)}{1+\beta(t) \sum\limits_{i=1}^{m} (x_i - t)^+} \quad .$$

The effect of censored observations is thus to lower the sample paths to the left of the censoring points while leaving the values to the right unchanged which appears inherently reasonable.

We next address ourselves to the question of the posterior distribution of $r(t)$ given exact observations. The answer to that question is given in the following theorem, i.e. that the posterior can be expressed as a continuous mixture of extended gamma distributions. However, the dimension of the mixing measure increases with sample size. The proof is given in section 7.

THEOREM 3.3 If the prior over the hazard rates is $\Gamma(\alpha(\cdot), \beta(\cdot))$ then the posterior over the hazard rates, given m observations of the form $X_1 = x_1, \ldots, X_m = x_m$ is a mixture of extended gamma processes. The distribution of the mixture is given by

$$(3.7) \qquad P(r(t) \in B | X_1 = x_1, \ldots, X_m = x_m) =$$

$$= \frac{\displaystyle\int_{[0,x_m)} \cdots \int_{[0,x_1)} \prod_{i=1}^{m} \hat{\beta}(z_i) \; F(B; \Gamma(\alpha + \sum_{i=1}^{m} I_{(z_i,\infty)}, \hat{\beta})) \prod_{i=1}^{m} d\left[\alpha + \sum_{j=i+1}^{m} I_{(z_j,\infty)}\right]^{(z_i)}}{\displaystyle\int_{[0,x_m)} \cdots \int_{[0,x_1)} \prod_{i=1}^{m} \hat{\beta}(z_i) \prod_{i=1}^{m} d\left[\alpha + \sum_{j=i+1}^{m} I_{(z_j,\infty)}\right](z_i)} \quad .$$

Here $F(B;Q)$ denotes the probability assigned to $B \in B_R$ by a stochastic process which is distributed as Q, $\hat{\beta}(\ )$ is defined as in (3.6), and the iterated integrations are done first with

respect to z_1 , then z_2 , through z_m . Of course

$$\sum_{j=m+1}^{m} I_{(z_j,\infty)}(z_i) \equiv 0 .$$

The complexity of this distribution makes it difficult to see how an observation $X_1 = x_1$ affects the posterior. Close examination reveals that a failure at time x_1 indicates an increase in the hazard rate prior to x_1 . However, this increase in the hazard rate diminishes as one looks further into the past. This is evidenced by the weight function $\hat{\beta}(t) = \beta(t)[1 + \beta(t)(x_1 - t)^+]^{-1}$ in the mixing integral. The above effect is tempered by the rate at which $\alpha(t)$ increases so that $\hat{\beta}(t)$ and $\alpha(t)$ together determine where and how the increase in risk (the unit jump in the α function) occurs.

4. DECREASING HAZARD RATES

With very little modification the work done for increasing hazard rates can be applied toward decreasing hazard rates. In particular, let $\alpha(\cdot)$, $\beta(\cdot)$, and $Z(\cdot)$ be defined as in section 2 with one exception. We assume that $\alpha(\cdot)$ and $\beta(\cdot)$ have finite values at plus infinity designated by $\alpha(\infty)$ and $\beta(\infty)$. We require that $\alpha(\infty) \geq \alpha(t)$, $t \geq 0$. $Z(\infty)$ is of course $G(\alpha(\infty), 1)$, and $Z(\infty) - \lim_{t\to\infty} Z(t)$ is independent of the rest of the process. We then define a decreasing extended gamma process $(D\Gamma(\alpha(\cdot), \beta(\cdot)))$ by

$$(4.1) \quad r(t) = \int_{[t,\infty)} \beta(s)dZ(s) + \beta(\infty)[Z(\infty) - \lim_{t\to\infty} Z(t)] = \int_{[t,\infty]} \beta(s)dZ(s) .$$

With this definition, $r(t)$ need not go to 0 as t goes to ∞. Integrals w.r.t. $\alpha(\cdot)$ are defined in an analogous manner. We take $r(t)$ to have non-increasing left-continuous paths. As expected,

$$
\text{E } r(t) = \int_{[t,\infty]} \beta(s) d\alpha(s),
$$

(4.2)

$$
\text{Var } r(t) = \int_{[t,\infty]} \beta^2(s) d\alpha(s), \quad \text{etc.}
$$

If one then uses a $D\Gamma(\alpha(t), \beta(t))$ prior over the failure rates, essentially all the distributional results of Section 3 carry over providing we replace "extended gamma" with "decreasing extended gamma", define $\hat{\beta}(\cdot)$ differently, and make our range of integration be $[t,\infty]$ rather than $[0,t)$. The following theorems will be stated without proofs.

THEOREM 4.1   If the prior over the hazard rates is $D\Gamma(\alpha(\cdot), \beta(\cdot))$, then the joint marginal survival function of n observations $X_1, \ldots, X_n$ is given by

(4.3) $\overline{F}(t_1,\ldots,t_n) = P(X_1 \geq t_1,\ldots,X_n \geq t_n) = \exp[-\int_{[0,\infty]} \ln(1+\beta(s) \sum_{i=1}^{n} \min(t_i,s)) d\alpha(s)].$

THEOREM 4.2   If the prior over the hazard rates is $D\Gamma(\alpha(\cdot), \beta(\cdot))$, then the posterior of the hazard rates given the n censored observations $X_1 \geq x_1, \ldots, X_n \geq x_n$ is $D\Gamma(\alpha(\cdot), \tilde{\beta}(\cdot))$ where

(4.4)

$$
\tilde{\beta}(t) = \frac{\beta(t)}{1+\beta(t) \sum_{i=1}^{n} \min(x_i,t)} .
$$

<u>THEOREM 4.3</u> If the prior over the hazard rates is $D\Gamma(\alpha(\cdot), \beta(\cdot))$, the posterior of the hazard rates given $X_1 = x_1, \ldots, X_m = x_m$ can be expressed as a continuous mixture of decreasing extended gamma distributions, i.e. as

$$(4.5) \qquad\qquad P(r(t)\epsilon B | X_1 = x_1, \ldots, X_m = x_m)$$

$$= \frac{\displaystyle\int_{[x_m,\infty]} \cdots \int_{[x_1,\infty]} \prod_{i=1}^{m} \tilde\beta(z_i) F(B; D\Gamma(\alpha + \sum_{i=1}^{m} I_{(z_i,\infty)}, \tilde\beta)) \prod_{i=1}^{m} d\left[\alpha + \sum_{j=i+1}^{m} I_{(z_j,\infty]}\right](z_i)}{\displaystyle\int_{[x_m,\infty]} \cdots \int_{[x_1,\infty]} \prod_{i=1}^{m} \tilde\beta(z_i) \prod_{i=1}^{m} d\left[\alpha + \sum_{j=i+1}^{m} I_{(z_j,\infty]}\right](z_i)}$$

The hazard rate estimation discussion in Section 5 will apply to the decreasing case provided one makes the obvious changes in the various expressions. Similarly, the computational results in Section 6 can easily be modified to handle the decreasing hazard rate situation.

5. <u>BAYES ESTIMATORS</u>

(a) Estimation of hazard rates.

A natural loss function to be used when estimating a hazard rate is the generalization of squared error loss given in Ferguson (1973). Thus our loss function will be

$$(5.1) \qquad\qquad L(r, \hat r) = \int_{[0,\infty)} (r(t) - \hat r(t))^2 dW(t)$$

where W is an artitrary finite measure on $[0,\infty)$ such that

$$\int_{[0,\infty)} \int_{[0,t)} \beta^2(s) d\alpha(s) dW(t) < \infty .$$

In finding $\hat{r}(t)$ which minimizes the expected loss, we may interchange the order of integration and thus minimize

$$E(r(t) - \hat{r}(t))^2$$

for a fixed t. The Bayes estimator is given by the posterior mean of $r(t)$.

If we ignore censored observations, we may use the form of $E\, r(t)$ in (2.3) and the fact that the mean of a mixture of distributions is the mixture of the means (assuming existence) to express the Bayes estimator of $r(t)$ as

$$(5.2) \quad \hat{r}(t) = \frac{\int_{[0,x_n)} \cdots \int_{[0,x_1)} \int_{[0,t)} \prod_{i=0}^{n} \hat{\beta}(z_i) \prod_{i=0}^{n} d[\alpha(z_i) + \sum_{j=i+1}^{n} I_{(z_j,\infty)}(z_i)]}{\int_{[0,x_n)} \cdots \int_{[0,x_1)} \prod_{i=1}^{n} \hat{\beta}(z_i) \prod_{i=1}^{n} d[\alpha(z_i) + \sum_{j=i+1}^{n} I_{(z_j,\infty)}(z_i)]}$$

where the iterated integrals are integrated with respect to z_0, z_1, z_2, etc., respectively.

Note that the denominator is of the exact same form as the numerator, though the integral is of a smaller dimension. Obviously $\hat{r}(t)$ is a nondecreasing function of t as expected. Including censored observations would only modify the $\hat{\beta}$ function.

While some approaches of nonparametric hazard rate estimation require the use of an arbitrary window function $w(\cdot)$ of integral one, this approach is free of any such function.

It would appear that the utility of this estimate is severely limited since it involves a multi-dimensional integral. We shall show in the next section, however, that $\hat{r}(t)$ is expressible in a manner that involves only one-dimensional integrals.

If the prime consideration is predictive in nature, the solution is different. Suppose

$$(5.3) \qquad \bar{F}*(t) = P(X_{n+1} \geq t \mid X_1 = x_1, \ldots, X_n = x_n)$$

denotes the conditional survival function of a future observation given n current observations. Then

$$(5.4) \qquad \bar{F}*(t) = \underset{r(t)|x_1,\ldots,x_n}{E} P(X_{n+1} \geq t \mid r(t), X_1=x_1,\ldots,X_n=x_n)$$

where the expectation is with respect to the posterior distribution of r given $X_1 = x_1, \ldots, X_n = x_n$. Since the X_i's are conditionally i.i.d., this is equivalent to

$$(5.5) \qquad \underset{r(t)|x_1,\ldots,x_n}{E} \exp\left[-\int_{[0,t)} r(s)\,ds\right]$$

which is the posterior mean at t of the random survival function $\bar{F}(t)$ defined from $r(t)$ by $\bar{F}(t) = \exp[-\int_{[0,t)} r(s)ds]$. Thus $\bar{F}*(t)$ can be thought of as the Bayes estimate of the survival function for the squared error loss function

$$(5.6) \qquad L(F,\hat{F}) = \int_{[0,\infty)} (F(t) - \hat{F}(t))^2 d\,W(t)$$

where $W(t)$ is a finite measure over $[0,\infty)$. Including censored

observations only changes the form of the $\hat{\beta}$ function.

To find a closed form expression for $\overline{F}^*(t)$, let

$$(5.7) \quad C = \int_{[0,x_n]} \cdots \int_{[0,x_1]} \prod_{i=1}^{n} \hat{\beta}(z_i) \prod_{i=1}^{n} d[\alpha(z_i) + \sum_{j=i+1}^{n} I_{(z_j,\infty)}(z_i)]$$

denote the norming constant in the posterior of $r(t)$. Since the posterior is a mixture of extended gamma distribution, we may use Theorem 3.1 to obtain

$$(5.8) \quad \overline{F}^*(t) =$$

$$= \frac{1}{C} \int_{[0,x_n]} \cdots \int_{[0,x_1]} \prod_{i=1}^{n} \hat{\beta}(z_i) \exp\left[-\int_{[0,\infty)} \ln(1+\hat{\beta}(z_0)(t-z_0)^+) d(\alpha(z_0) + \sum_{i=1}^{n} I_{(z_i,\infty)}(z_0))\right] \prod_{i=1}^{n} d[\alpha(z_i) + \sum_{j=i+1}^{n} I_{(z_j,\infty)}(z_i)]$$

The integrand can be evaluated as

$$(5.9) \quad \prod_{i=1}^{n} \hat{\beta}(z_i) \exp\left[-\int_{[0,\infty)} \ln(1+\hat{\beta}(z_0)(t-z_0)^+) d\alpha(z_0) - \sum_{i=1}^{n} \ln(1+\hat{\beta}(z_i)(t-z_i)^+\right]$$

$$= \exp\left[-\int_{[0,\infty)} \ln(1+\hat{\beta}(z_0)(t-z_0)^+) d\alpha(z_0)\right] \prod_{i=1}^{n} \frac{\hat{\beta}(z_i)}{1+\hat{\beta}(z_i)(t-z_i)^+} .$$

Noting that our first factor is free of z_i, and relabeling

$$(5.10) \quad \beta^*(z_i) = \frac{\hat{\beta}(z_i)}{1+\hat{\beta}(z_i)(t-z_i)^+} ,$$

we obtain

(5.11) $$\bar{F}^*(t) = \exp\left[-\int_{[0,\infty)} \ln(1+\hat{\beta}(z_0)(t-z_0)^+)d\alpha(z_0)\right]$$

$$\cdot\ \frac{\displaystyle\int_{[0,x_n)}\cdots\int_{[0,x_1)}\prod_{i=1}^{n}\beta^*(z_i)\prod_{i=1}^{n}d[\alpha(z_i)+\sum_{j=i+1}^{n}I_{(z_j,\infty)}(z_i)]}{\displaystyle\int_{[0,x_n)}\cdots\int_{[0,x_1)}\prod_{i=1}^{n}\hat{\beta}(z_i)\prod_{i=1}^{n}d[\alpha(z_i)+\sum_{j=i+1}^{n}I_{(z_j,\infty)}(z_i)]}\ \cdot$$

Similarly, Corollary 3.1 can be used to express the joint survival function of k future observations $X_{n+1}, \ldots, X_{n+k}$ conditional on the observed data. Thus

(5.12) $$\bar{F}^*(t_{n+1},\ldots,t_{n+k}) = P(X_{n+1} \geq t_{n+1},\ldots,X_{n+k} \geq t_{n+k} \mid X_1 = x_1,\ldots,X_n = x_n)$$

is of the same form as (5.11) with $(t-z_i)^+$ replaced by $\sum_{j=n+1}^{n+k}(t_j - z_i)^+$. One of the consequences of this is that the minimum of k future observations has the conditional survival function given in (5.11) with $\hat{\beta}$ replaced by $k\hat{\beta}$.

Noting in (5.11) that β^* is a nonincreasing function of t which is equal to $\hat{\beta}$ when $x = 0$ guarantees that $\bar{F}^*(t)$ is a bonafide survival function. The first factor of $\bar{F}^*(t)$ in (5.11) would be the survival function of a future observation were the observations censored at $x_1,\ldots,x_n$ rather than observed. Thus the second factor contains the information gained by observing "deaths" rather than "losses" (see Kaplan and Meier (1958) for elaboration on this terminology).

Note that $\overline{F}^*(t)$ is differentiable. By using the product rule for derivatives, interchanging derivatives and integrals, and interchanging the order of integration, it can be shown that the density corresponding to $\overline{F}^*(t)$ is given by

(5.13)
$$f^*(t) = \exp[-\int_{[0,\infty)} \ln(1 + \hat{\beta}(t-z_0)^+)d\alpha(z_0) \cdot$$

$$\cdot \frac{\int_{[0,x_n)} \cdots \int_{[0,x_1)} \int_{[0,x)} \prod_{i=0}^{n} \beta^*(z_i) \prod_{i=0}^{n} d[\alpha(z_i) + \sum_{j=i+1}^{n} I_{(z_j,\infty)}(z_i)]}{\int_{[0,x_n)} \cdots \int_{[0,x_1)} \prod_{i=1}^{n} \hat{\beta}(z_i) \prod_{i=1}^{n} d[\alpha(z_i) + \sum_{j=i+1}^{n} I_{(z_j,\infty)}(z_i)]} .$$

Moreover, if we define the random density function by

(5.14)
$$f(t) = r(t)\exp[-\int_{[0,t)} r(s)ds] = -\frac{d}{dt}\overline{F}(t) ,$$

then by interchanging differentiation and integration over the posterior distribution we have

(5.15)
$$f^*(t) = -\frac{d}{dt}E(\overline{F}(t)) = E(-\frac{d}{dt}\overline{F}(t)) = E(f(t))$$

so that $f^*(t)$ is the Bayes estimate of the density with the usual type loss function,

(5.16)
$$L(f,\hat{f}) = \int_{[0,t)} (f(t) - \hat{f}(t))^2 dW(t) .$$

This suggests an approach to density estimation which gives smooth continuous estimates and avoids the problem of defining window functions as in Rosenblatt (1971).

Finally , we may obtain the failure rate $r^*(t)$ corresponding to $\overline{F}^*(t)$ as

$$(5.17) \qquad r^*(t) = \frac{f^*(t)}{\overline{F}^*(t)} =$$

$$= \frac{\int\limits_{[0,x_n)} \cdots \int\limits_{[0,x_1)} \int\limits_{[0,t)} \prod_{i=0}^{n} \beta^*(z_i) \prod_{i=0}^{n} [d\alpha(z_i) + \sum_{j=i+1}^{n} I_{[z_j,\infty)}(z_i)]}{\int\limits_{[0,x_n)} \cdots \int\limits_{[0,x_1)} \prod_{i=1}^{n} \beta^*(z_i) \prod_{i=1}^{n} [d\alpha(z_i) + \sum_{j=i+1}^{n} I_{[z_j,\infty)}(z_i)]}$$

However, this is the same expression as in (5.2) with the exception that the $\hat{\beta}$'s are replaced by β^*'s. In other words, the effect of using the loss function over the c.d.f.'s (5.6) when estimating a distribution (be it c.d.f., density, or hazard rate) at a point t rather than the loss function over the hazard rate (5.1) is merely to act as though one has an additional censored observation at the point t .

If one is interested in estimating the mean of the distribution in question, then

$$(5.18) \qquad \mu = \int\limits_{[0,\infty)} \overline{F}(t)dt \equiv \int\limits_{[0,\infty)} \exp[- \int\limits_{[0,t)} r(s)ds] \, dt$$

is a well defined random variable providing

$$\int_{[0,\infty)} \exp\left[-\int_{[0,t)} \ln\left(1+\beta(s)(x-s)\right)d\alpha(s)\right]dt < \infty \quad .$$

Taking expectations with respect to the posterior distribution, the mean of the estimated survival function $\overline{F}^*$,

$$(5.19) \qquad \mu^* = \int_{[0,\infty)} \overline{F}^*(t)dt = \int_{[0,\infty)} E(\overline{F}(t))dt = E(\mu)$$

is the Bayes estimate of μ under the usual loss function

$$(5.20) \qquad L(\mu,\hat{\mu}) = (\mu-\hat{\mu})^2 \quad .$$

6.   COMPUTATION AND SIMULATION.

The presence of the multi-dimensional integral which occurs in our estimates would appear to make computation extremely difficult. The following theorem enables us to work with integrals of only one dimension. The integrands are powers of the $\hat{\beta}$ function and the integration is with respect to the α measure.

__THEOREM 6.1__   Assuming that $\alpha(\cdot)$ and $\beta(\cdot)$ are defined as in Section 2, then

$$
(6.1) \qquad \int_{[0,x_n)} \cdots \int_{[0,x_1)} \prod_{i=1}^{n} \hat{\beta}(z_i) \prod_{i=1}^{n} d[\alpha(z_i) + \sum_{j=i+1}^{n} I_{(z_j,\infty)}(z_i)]
$$

$$
= \sum_{\underset{\sim}{e}} k(\underset{\sim}{e}) [\prod_{\{i,e_i \geq 1\}} \int_{[0,x_i)} \hat{\beta}(t)^{e_i} d\alpha(t)]
$$

where $0 < x_n \leq x_{n-1} \leq \ldots \leq x_1 < \infty$, the sum is over all vectors $\underset{\sim}{e} = (e_1, \ldots, e_n)$ of non-negative integers such that

$$
\sum_{i=1}^{j} e_i \leq j, \; j = 1, \ldots, n-1; \; \sum_{i=1}^{n} e_i = n; \quad \text{and}
$$

$$
(6.2)
$$
$$
k(\underset{\sim}{e}) = \prod_{\{j; e_j \geq 2\}} [(j-1) - \sum_{1}^{j-1} e_i] P_{e_j - 1} = \prod_{\{j, e_j \geq 2\}} [(j-1) - \sum_{1}^{j-1} e_i]! / [j - \sum_{1}^{j} e_j]!
$$

where nP_r denotes the number of permutations of n things taken r at a time.

PROOF: Since the inside integral can be expressed as a sum of n integrals, the next as a sum of $n-1$ integrals, etc., it is clear that (6.1) can be expressed as a sum of $n!$ integrals. Moreover, since we assume the x_i's are ordered in decreasing fashion, it must be the case that

$$(6.3) \quad \int_{[0,x_i)} \hat{\beta}(z_i)\, dI_{(z_j,\infty)}(z_i) = \hat{\beta}(z_j) \quad \text{for} \quad z_j < x_j \leq x_i,\ j > i.$$

This will then combine with $\hat{\beta}(z_j)^k$ to give $\hat{\beta}(z_j)^{k+1}$ for some integer k. Close scrutiny will reveal however that the exponent k of $\hat{\beta}(z_j)^k$ can never exceed j. Moreover, each of the $n!$ integrals will be of the form

$$(6.4) \quad \prod_{\{i,e_i \geq i\}} \int_{[0,x_i)} \hat{\beta}(t)^{e_i}\, d\alpha(t)$$

where $\sum_{i=1}^{n} e_i = n$, the e_i being non-negative integers. Thus to establish Theorem 7.1, we need only argue that $k(\underset{\sim}{e})$ correctly counts the number of terms of the form given in (6.4).

Consider a vector $\underset{\sim}{e} = (e_1, \ldots, e_n)$ of the form specified in the statement of the theorem. Fix j and assume that $e_j \geq 1$. Then one unit of the exponent of $\hat{\beta}(z_j)^{e_j}$ must come from the jth integration, and the other $e_j - 1$ units must come from previous integrations. Since there will be $(j-1) - \sum_{i=1}^{j-1} e_i$ previous integrations unaccounted for, there are

$$(6.5) \quad \binom{(j-1) - \sum_{i=1}^{j-1} e_i}{e_j - 1}$$

ways of choosing the required $e_j - 1$ integrations. Moreover, the first chosen integration can increase the exponent in $e_j - 1$ ways (by being routed to any of the other integrations which eventually contribute to e_j), the second chosen integration can increase the exponent in $e_j - 2$ ways, etc. Thus we need to multiply (6.5) by $(e_j - 1)!$ to count how many ways we can obtain the necessary exponent. Using the multiplication principle then to count the total number of terms (6.4) for a given vector $\underset{\sim}{e}$ gives us $k(\underset{\sim}{e})$.

Example Consider the very specialized case where $\alpha(\cdot)$ jumps at 0 and is then flat. That is

$$\alpha(0) = 0$$
$$\alpha(x) = \alpha, \ x > 0.$$

In this case, $r(t)$ will be a constant function whose value will be a $G(\alpha, \beta(0))$ random variable. Thus the only value of $\beta(t)$ that matters is $\beta(0) = \beta$. Since the parameter in an exponential distribution is just its constant failure rate, this is equivalent to putting a $G(\alpha, \beta)$ prior over the parameter θ of an exponential density.

Then if we have complete observations at $x_1, \ldots, x_n$ and censored observations at $x_{n+1}, \ldots, x_{n+m}$, we can specify our posterior distribution of $r(t)$ from Theorems 3.2 and 3.3. Since the posterior of an exponential distribution with a gamma prior is again a gamma, the distribution of $r(t_0)$, $t_0 > 0$ specified by the mixture in Theorem 3.3 must also be a gamma distribution.

In this event, the Bayes estimate of $r(t)$, $t > 0$, is a

constant (free of t) and may be expressed in terms of Theorem 6.1.

Let $\#\underset{\sim}{e}$ denote the number of non-zero components of $\underset{\sim}{e}$. Then from Theorem 6.1, the numerator of $\hat{r}(t)$ equals

$$\hat{\beta}(0)^{n+1} \sum_{\underset{\sim}{e}} k(\underset{\sim}{e}) \; \alpha^{\#\underset{\sim}{e}}$$

$$= \hat{\beta}(0)^{n+1} \sum_{i=1}^{n+1} \alpha^i \sum_{\{\underset{\sim}{e};\#\underset{\sim}{e}=i\}} k(\underset{\sim}{e}) \; .$$

However, it can be shown that $\sum_{\{\underset{\sim}{e};\#\underset{\sim}{e}=i\}} k(\underset{\sim}{e})$ is the coefficients of Z^i in $Z(Z+1)(Z+2)\ldots(Z+n)$ (the modulus of Sterling numbers of the first kind). Thus the numerator of $\hat{r}(t)$ equals

$$\left(\frac{\beta}{1+\beta \sum\limits_{i=1}^{n+m} x_i} \right)^{n+1} \alpha(\alpha+1)\ldots(\alpha+n) \; .$$

By similar treatment, the denominator of $\hat{r}(t)$ equals

$$\left(\frac{\beta}{1+\beta \sum\limits_{1}^{n+m} X_i} \right)^{n} \alpha(\alpha+1)\ldots(\alpha+n-1).$$

Thus $\hat{r}(t) = \left(\dfrac{\beta}{1+\beta \sum x_i} \right) (\alpha+n) = \dfrac{(\alpha+n)}{\left(\dfrac{1}{\beta} + \sum\limits_{1}^{n+m} x_i \right)}$, $t > 0$.

This agrees with the posterior mean for uncensored data given in Mann, Schafer, and Singpurwalla (1974) (see page 414). As one would expect, as $n \to \infty$,

$$\hat{r}(t) \sim [\text{total no. of failures}]/[\text{total time on test}].$$

In order to observe the performance of our Bayes estimators, samples from Weibull and exponential distributions were taken and the corresponding Bayes estimators computed. In all cases the sample size was 11 and the prior parameter functions $\alpha(t) = t$, and $\beta(t) \equiv 2$ were used. Thus the expected value of the prior hazard rate would be $\int_{[0,t)} \beta(s)\,d\alpha(s) = 2t$. This is the hazard rate of a Weibull distribution with mean .8862. All observations were complete (not censored). It is true that if one decreases $\beta(\cdot)$ and increases $\alpha(\cdot)$ in such a way that the mean of the prior $\int_{[0,t)} \beta(s)\,d\alpha(s)$ is unchanged, the variance of the prior $\int_{[0,t)} \beta^2(s)\,d\alpha(s)$ will be decreased. This has the effect of specifying a more precise prior distribution and hence the prior will have more influence in posterior estimates.

Bayes estimates are computed under both loss functions i.e. squared error loss on hazard rates and c.d.f.'s. The hazard rate corresponding to the Bayes estimate of the c.d.f. is graphed along with the estimated hazard rate on the hazard rate graphs for the purpose of comparison. Similarly, the c.d.f. corresponding to the Bayes estimate of the hazard rate is graphed on the c.d.f. graphs. Thus on figures 1-6, the posterior Bayes estimate of the hazard rate is denoted by a solid line, while the hazard rate which corresponds to the posterior Bayes estimate of the c.d.f. is denoted by the line made up of alternate dashes and plusses. Since the key is the same for all graphs, it is stated explicitly only in Figure 1. A similar interpretation holds for figures 7-12 concerning the c.d.f.'s. Thus

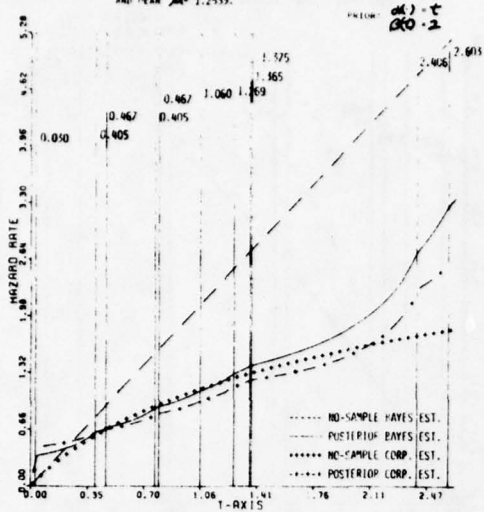FIGURE 1. SAMPLE FROM WEIBULL DISTRIBUTION WITH HAZARD RATE $r(t) = t$ AND MEAN $\mu = 1.2533$. PRIOR: $\alpha(t) = t$, $\beta(t) = 2$

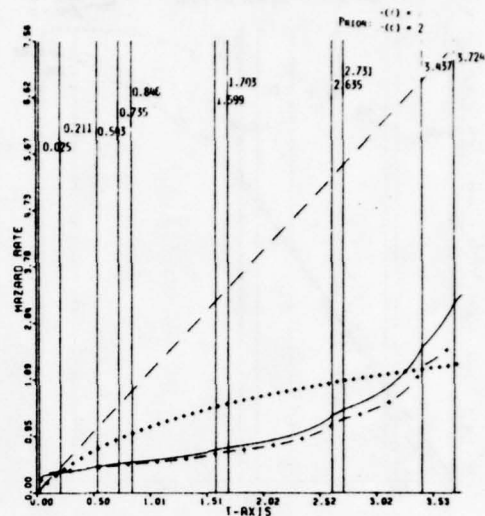FIGURE 2. SAMPLE FROM EXPONENTIAL DISTRIBUTION WITH HAZARD RATE $r(t) = .798$ AND MEAN $\mu = 1.2533$. PRIOR: $\alpha(t) = t$, $\beta(t) = 2$

FIGURE 3. SAMPLE FROM WEIBULL DISTRIBUTION WITH HAZARD RATE $r(t) = 2t$ AND MEAN $\mu = .8862$. PRIOR: $\alpha(t) = t$, $\beta(t) = 2$
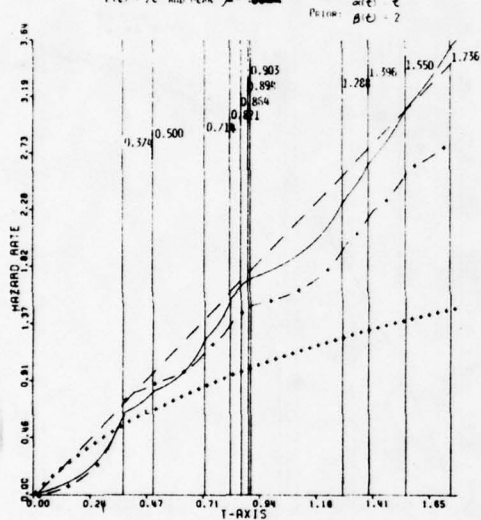
FIGURE 4. SAMPLE FROM EXPONENTIAL DISTRIBUTION WITH HAZARD RATE $r(t) = 1.128$ AND MEAN $\mu = .8862$. PRIOR: $\alpha(t) = t$, $\beta(t) = 2$
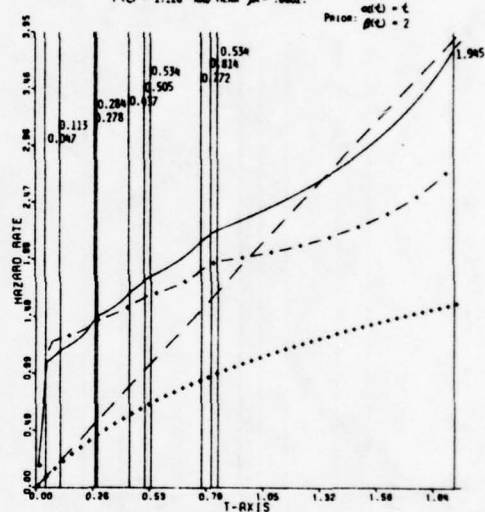
FIGURE 5. SAMPLE FROM WEIBULL DISTRIBUTION WITH HAZARD RATE $r(t) = 3t$ AND MEAN $\mu = .7236$. PRIOR: $\alpha(t) = t$, $\beta(t) = 2$
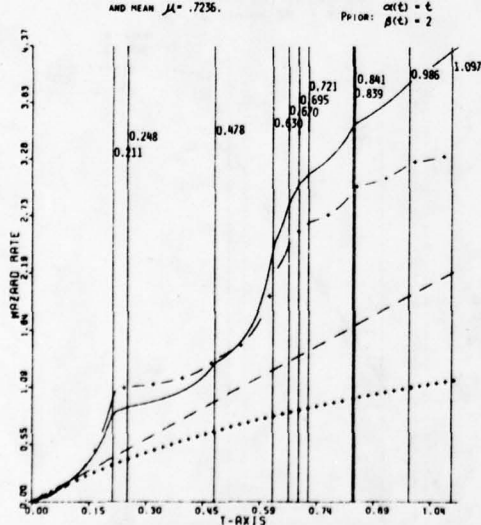
FIGURE 6. SAMPLE FROM EXPONENTIAL DISTRIBUTION WITH HAZARD RATE $r(t) = 1.382$ AND MEAN $\mu = .7236$. PRIOR: $\alpha(t) = t$, $\beta(t) = 2$
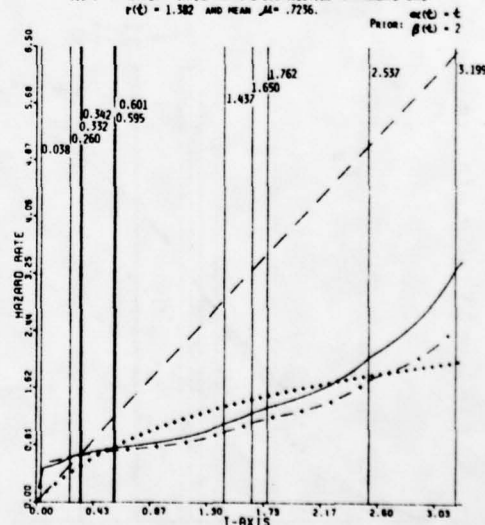
Legend:
- .... NO-SAMPLE BAYES EST.
- —·—·— POSTERIOR BAYES EST.
- ••••••• NO-SAMPLE CORR. EST.
- ···•··· POSTERIOR CORR. EST.

FIGURE 7. Sample from Weibull distribution with hazard rate $\rho(t) = t$ and mean $\mu = 1.2533$. Prior: $\alpha(t) = t$, $\beta(t) = 2$.
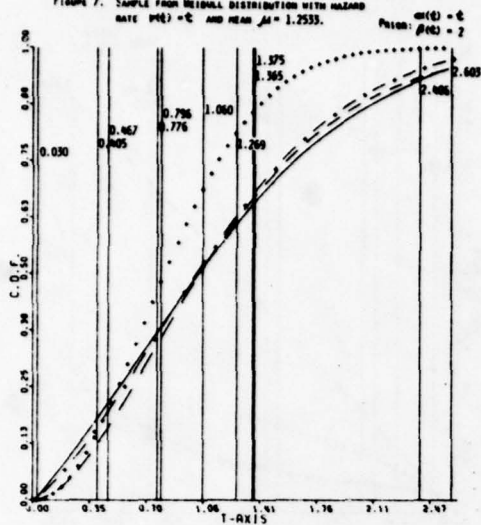
FIGURE 8. Sample from exponential distribution with hazard rate $r(t) = .798$ and mean $\mu = 1.2533$. Prior: $\alpha(t) = t$, $\beta(t) = 2$.
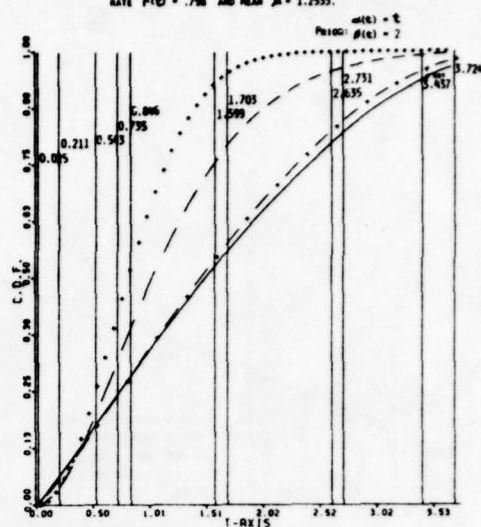

FIGURE 9. Sample from Weibull distribution with hazard rate $\rho(t) = 2t$ and mean $\mu = .8862$. Prior: $\alpha(t) = t$, $\beta(t) = 2$.
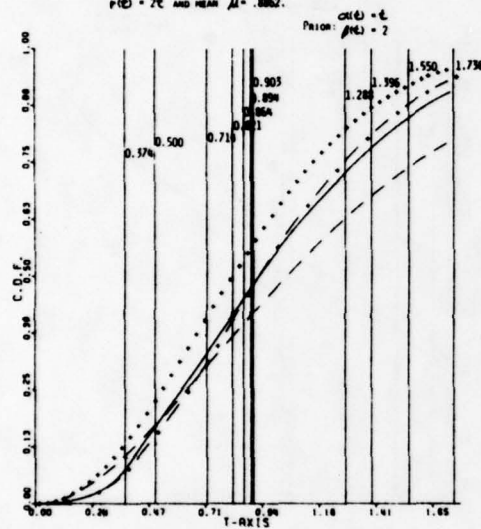
FIGURE 10. Sample from exponential distribution with hazard rate $r(t) = 1.128$ and mean $\mu = .8862$. Prior: $\alpha(t) = t$, $\beta(t) = 2$.
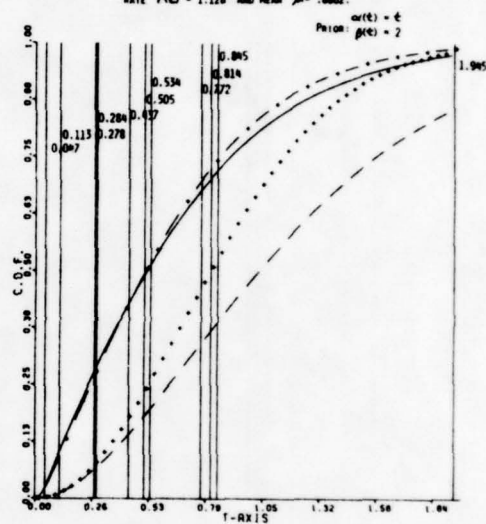

FIGURE 11. Sample from Weibull distribution with hazard rate $r(t) = 3t$ and mean $\mu = .7236$. Prior: $\alpha(t) = t$, $\beta(t) = 2$.
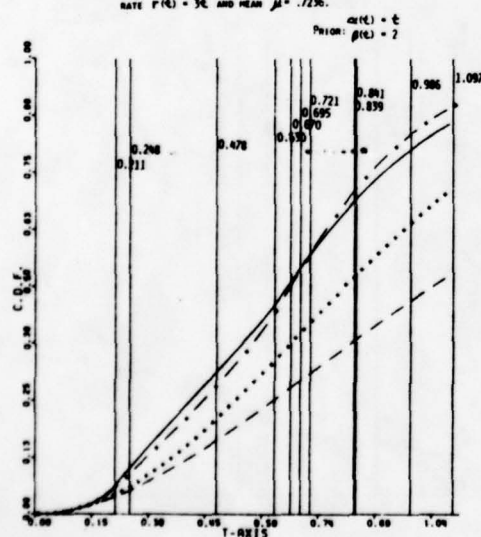
FIGURE 12. Sample from exponential distribution with hazard rate $r(t) = 1.382$ and mean $\mu = .7236$. Prior: $\alpha(t) = t$, $\beta(t) = 2$.
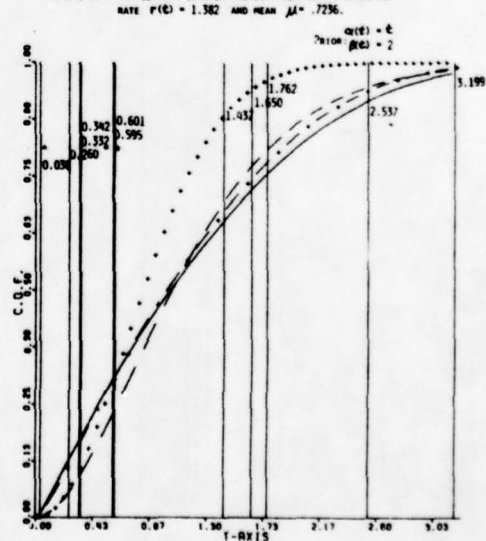
Figure 13. Data from Kaplan-Meier paper. (Starred lines indicate censored data.)

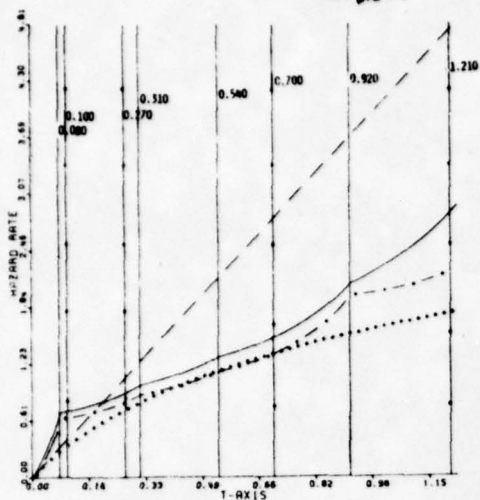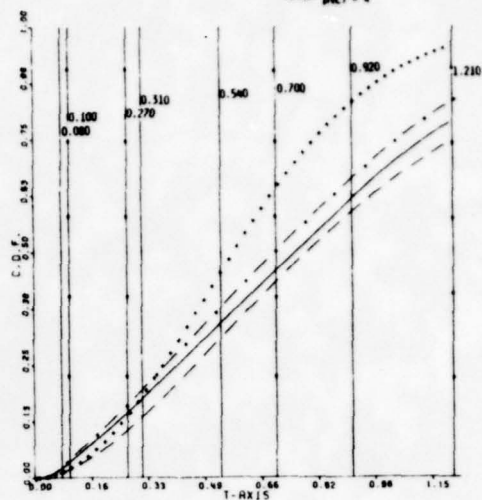Prior: $\alpha(t) = t$
$\beta(t) = 4$

Figure 14. Data from Kaplan-Meier paper. (Starred lines indicate censored data.)

Prior: $\alpha(t) = t$
$\beta(t) = 4$

the distributions corresponding to the solid (plus-dash) lines in
figures 1-6 are the same as the distributions corresponding to the
alternating plus-dash (solid) lines in figures 7-12 respectively.
Figures 1, 3, and 5 depict the Bayes estimates of the hazard rates
when the random samples come from Weibull distributions whose failure
rates are respectively t, 2t, and 3t. Note that the estimates reflect
the populations from which the samples come by generally having
progressively steeper slopes. (Note that the scales change between
graphs so that visual slopes are deceptive.) Figures 2, 4, 6, depict the
Bayes estimates of the hazard rates when the samples come from
exponential distributions. In each case the mean of the exponential
is made to be the same as the previous Weibull distribution.

The purpose of this is to see if our estimated hazard rates
will reflect the difference between Weibull and exponential distri-
butions. Note that in each case, the estimated hazard rates are
flatter for the exponential distributions than for the Weibull
distributions. Of course since exponential distributions are on
the boundary (our prior puts all the probability on nondecreasing
hazard rates), our estimates of the hazard rate will necessarily
be increasing to some degree. Figures 7-12 give the estimates in
terms of c.d.f.'s rather than hazard rates. They are of course
based on the same samples used in Figures 1-6. Finally, Figure 13
and 14 depict estimates of the hazard rate and c.d.f. for the data
given in the Kaplan and Meier (1958) paper. The prior was
arbitrarily taken to be $\alpha(t) = t$, $\beta(t) \equiv 4$ and the starred lines

indicate cencored values. Note that a slight peaking occurs in estimates of the hazard rate at complete observations, although this peaking is scarcely detectable in the c.d.f.'s. In comparing these graphs with the estimates of the c.d.f given in the papers by Susarla and Van Ryzin and Ferguson and Phadia, it appears that our estimate is closer to the Kaplan-Meier product limit estimate than theirs. We feel that our continuous estimates are more appealing than their discontinuous estimates.

In conclusion, it appears that our estimates have the property of being smooth and continuous and yet are very responsive to the data.

7. PROOFS OF THEOREMS

In this section we take our stochastic processes to be defined on an arbitrary probability space (Ω, F, P). We use R^R to denote the set of all nonnegative functions on the nonnegative real line R and B_R to denote the smallest σ-algebra generated by sets of the form $\{x(\cdot) \epsilon R^R : x(t_1) \epsilon I_1, \ldots, x(t_k) \epsilon I_k\}$ where $I_1, \ldots, I_k$ are intervals in R. A stochastic process r is a measurable function which maps Ω into R^R. This induces a probability measure on (R^R, B_R) called the distribution of the process r. Since, with probability one, the sample paths $r(t,\omega)$ of our stochastic process are failure rates, we can define a probability measure $\tilde{P}$ on the product space $(R^R \times R, B_R \times B)$ by extending $\tilde{P}(B \times C) = \int_A F_\omega(C) \, dP(\omega)$ to the usual product σ-algebra of B_R and the Borel sets B. Here $A = r^{-1}(B)$ and $F_\omega(C)$ is the probability assigned

to C by the distribution corresponding to $r(\cdot,\omega)$. Then a probability measure on $(R, \mathcal{B})$ is determined by $\hat{P}(C) = \tilde{P}(R^R \times C) = \int_\Omega F_\omega(C)\, dP(\omega)\ \forall\, C \in \mathcal{B}$. The posterior distribution of the process for a single observation is a function $\phi(\cdot, \cdot) : \mathcal{B}_R \times R \rightarrow [0, 1]$ Borel measurable in the second argument when the first argument is fixed such that for each fixed $x \in R$, $\phi(\cdot, x)$ is a probability measure on $(R^R, \mathcal{B}_R)$ and $\int_C \phi(B, x)\, d\hat{P}(x) = \tilde{P}(B \times C)$ for all $B \in \mathcal{B}_R$ and $C \in \mathcal{B}$. The extension for several observations is straightforward.

For convenience in writing we adopt the following notation:

(i) $g(x; \alpha,\beta) = x^{\alpha-1}\exp(-x/\beta)I_{[0,\infty)}(x)/\Gamma(\alpha)\alpha^\beta$; $g(x;\alpha) = g(x;\alpha,1)$.

(ii) $\Delta\alpha_i = \alpha(t_i^{(n)})-\alpha(t_{i-1}^{(n)})$, $i = 1, \ldots, k(n)$

(iii) $\beta_i = \beta(t_i^{(n)})$, $i = 1, \ldots, k(n)$

(iv) $\Sigma = \sum_{i=1}^{k(n)}$ and $\Pi = \prod_{i=1}^{k(n)}$

(v) $B_n(u,\beta,\tau,y) = \{(u_1,\ldots,u_{k(n)}) \in R^{k(n)} : \sum_{t_i<\tau_1} \beta_i u_i > y_1, \sum_{\tau_1 \le t_i < \tau_2} \beta_i u_i > y_2, \ldots, \sum_{\tau_{k-1} \le t_i < \tau_k} \beta_i u_i > y_k\}$. Often $B_n(u,\beta,\tau,y)$ is abbreviated $B_n(u,\beta)$.

(vi) $F(B;Q) =$ Probability assigned to $B \in \mathcal{B}_R$ by a stochastic process with distribution Q.

LEMMA 7.1

Let $\alpha(\cdot)$ be a nonnegative nondecreasing left continuous function on $[0, \infty)$ with $\alpha(0) = 0$. For a sequence of partitions $0 = t_0 < t_1 < \ldots < t_{k(n)} < \infty$ whose norm goes to zero and whose

upper end point goes to infinity, define $a_n(0) = 0$ and

$$\alpha_n(t) = \sum_{i=1}^{k(n)} \alpha(t_i) I_{(t_{i-1}, t_i]}(t) + \alpha(t_{k(n)}) I_{(t_{k(n)}, \infty)}(t), \quad t \varepsilon (0, \infty).$$

If we define B and $r_n(t, \omega)$ as in (7.6) and (2.5) respectively, $A = r^{-1}(B)$, and $A_n = r_n^{-1}(B)$, then

$$(7.1) \qquad \int_\Omega I_{A_n}(\omega) \, dP(\omega) \longrightarrow \int_\Omega I_A(\omega) \, dP(\omega)$$

i.e. $F(B; \Gamma(\alpha_n, \beta)) \longrightarrow F(B; \Gamma(\alpha, \beta))$ and

$$(7.2) \qquad \lim_{n \to \infty} \int_{B_n(\underline{u}, \beta)} \Pi \, g(u_i; \Delta\alpha_i) \, du_i = F(B; \Gamma(\alpha, \beta)).$$

PROOF: For almost all ω, $r_n(\tau, \omega) \to r(\tau, \omega)$ uniformly on $[0, t]$, $0 < t < \infty$ since $r_n(\tau, \omega)$ and $r(\tau, \omega)$ are almost surely non-decreasing left continuous bounded functions on $\tau \varepsilon [0, t]$. Thus $I_{A_n}(\omega) \xrightarrow{a.s.} I_A(\omega)$ and (7.1) follows by LDCT. Note that

$$(7.3) \quad \int_\Omega I_{A_n}(\omega) \, dP(\omega) = F(B; \Gamma(\alpha_n, \beta)) = \int_{B_n(u, \beta)} \Pi \, g(u_i; \Delta\alpha_i) \, du_i.$$

Thus the l.h.s. of (7.2) exists and is equal to $F(B; \Gamma(\alpha, \beta))$.

PROOF OF THEOREM 3.1: Define $r_n(\tau)$ as in (2.5). As noted before, $Z(t, \omega)$ is nondecreasing and left continuous for almost all ω. Hence $r(\tau)$, $\tau \varepsilon [0, t]$ is bounded almost surely and $r_n(\tau) \to r(\tau)$ for each $\tau \varepsilon [0, t]$. Thus $\int_{[0, t]} r_n(\tau) \, d\tau \longrightarrow \int_{[0, t]} r(\tau) \, d\tau$ (by LDCT) for almost all $\omega \varepsilon \Omega$.

Now, $P(X \geq t) = \int_\Omega P(X \geq t | r(\cdot, \omega)) dP(\omega)$

$$= \int_\Omega \exp[- \int_{[0,t)} r(\tau) \, d\tau] dP(\omega)$$

$$= \lim_{n \to \infty} \int_\Omega \exp[- \int_{[0,t)} r_n(\tau) d\tau] dP(\omega) \quad \text{by LDCT}$$

$$= \lim_{n \to \infty} \int_\Omega \exp[-\Sigma(t-t_i)^+ \beta_i \{Z(t_i) - Z(t_{i-1})\}] dP(\omega)$$

$$= \lim_{n \to \infty} \int_{R^{k(n)}} \exp[-\Sigma(t-t_i)^+ \beta_i u_i] \Pi \, g(u_i; \Delta\alpha_i) du_i$$

(7.4) $$= \lim_{n \to \infty} \Pi (1 + (t-t_i)^+ \beta_i)^{-\Delta\alpha_i}$$

(7.5) $$= \exp[- \int_{[0,t)} \ln(1 + \beta(s)(t-s)) d\alpha(s)].$$

PROOF OF THEOREM 3.2: First consider the case $m = 1$. Define $B \in \mathcal{B}_R$ by

(7.6) $\quad B = \{r(\cdot) \in R^R: r(\tau_1) > y_1, r(\tau_2) - r(\tau_1) > y_2, \ldots, r(\tau_k) - r(\tau_{k-1}) > y_k\}$

where k is an arbitrary positive integer and $\tau_1 < \cdots < \tau_k$, $y_1, \ldots, y_k$ are arbitrary nonnegative real numbers. It can be shown that the distribution of the process $r(t, \omega)$ is uniquely determined by the probabilities given to sets of the form $A = r^{-1}(B)$. Thus it suffices to show that the posterior probability of sets of the form $A = r^{-1}(B)$ equals that assigned by an extended gamma process with parameter functions $\alpha(\cdot)$ and $\hat{\beta}(\cdot)$.

Defining $r_n(t)$ as in (2.5) and $A_n = r_n^{-1}(B)$, then $r_n(t) \xrightarrow{a.s.} r(t)$ and $I_{A_n} \xrightarrow{a.s.} I_A$. Thus

$$P(r(t)\epsilon B \mid X \geq x) = P(r(t)\epsilon B, X \geq x)/P(X \geq x)$$

$$= \int_A \exp[-\int_{[0,x)} r(t)dt]dP(\omega)/\int_\Omega \exp[-\int_{[0,x)} r(t)dt]dP(\omega)$$

$$= \lim_{n\to\infty} \int_\Omega \exp[-\int_{[0,x)} r_n(t)dt]I_{A_n}(\omega)dP(\omega)/\lim_{n\to\infty} \int_\Omega \exp[-\int_{[0,x)} r_n(t)dt]dP(\omega)$$

$$= \lim_{n\to\infty}\int_{B_n(\underset{\sim}{u},\beta)} \exp[-\Sigma\beta_i(x-t_i)^+ u_i]\Pi g(u_i,\Delta\alpha_i)du_i/\lim_{n\to\infty}\Pi[1+\beta_i(x-t_i)^+]^{-\Delta\alpha_i} \quad \text{by}(7.4)$$

$$= \lim_{n\to\infty} \int_{B_n(\underset{\sim}{u},\beta)} \Pi g(u_i;\Delta\alpha_i, 1+\beta_i(x-t_i)^+)du_i$$

$$= \lim_{n\to\infty} \int_{B_n(\underset{\sim}{v},\hat\beta)} \Pi g(v_i;\Delta\alpha_i)dv_i \quad \text{where} \quad \beta_i u_i = \hat\beta_i v_i$$

$$= F(B;\Gamma(\alpha,\hat\beta)) \quad \text{by (7.2) of Lemma 7.1.}$$

This proves the theorem for $m = 1$. Using a parenthesized subscript to emphasize explicitly the dependence of $\hat\beta$ on the sample size, we have $\hat\beta_{(j)}(t)/[1+\hat\beta_{(j)}(t)(x_{j+1}-t)^+] = \hat\beta_{(j+1)}(t)$. The theorem follows by induction.

PROOF OF THEOREM 3.3: First consider the case $m = 1$. It suffices to consider sets of the form B where B is defined by (7.6) and show that

$$(7.7) \qquad \int_{[x,\infty)} \phi(B,x) f(x) dx = P(r(t)\epsilon B, X \geq x)$$

where

$$(7.8) \qquad f(x) = -\frac{d}{dx} \bar F(x) = -\frac{d}{dx} \int_\Omega \exp[-\int_{[0,x)} r(s)ds]dP(\omega)$$

is the marginal density of X and

(7.9) $\quad \phi(B,x) = \int\limits_{[0,x)} \hat{\beta}(s)F(B;\Gamma(\alpha+I_{(s,\infty)},\hat{\beta}))d\alpha(s) / \int\limits_{[0,x)} \hat{\beta}(s)d\alpha(s)$

is the conditional distribution of the process. To show (7.7) we define $\phi_n(B,x)$ and $f_n(x)$ below by (7.11) and (7.12) respectively and prove the following series of claims:

<u>Claim 1</u>: (7.10) $\int\limits_{[x,\infty)} \phi_n(B,x)f_n(x)dx = \int\limits_{\Omega} \exp[-\int\limits_{[0,x)} r_n(s)ds]I_{A_n}(\omega)dP(\omega)$

where r_n is defined by (2.5) and $A_n = r_n^{-1}(B)$.

<u>Claim 2</u>: The r.h.s. of (7.10) converges to $P(r(t)\varepsilon B, X \geq x)$ as $n \to \infty$.

<u>Claim 3</u>: The l.h.s. of (7.10) converges to $\int\limits_{[x,\infty)} \phi(B,x)f(x)dx$ as $n \to \infty$.

We define

(7.11) $\phi_n(B,x) = \dfrac{d}{dx}\int\limits_{\Omega} \exp[-\int\limits_{[0,x)} r_n(s)ds]I_{A_n}(\omega)dP(\omega) / \dfrac{d}{dx}\int\limits_{\Omega}\exp[-\int\limits_{[0,x)} r_n(s)ds]dP(\omega)$

and

(7.12) $\quad f_n(x) = -\dfrac{d}{dx}\int\limits_{\Omega} \exp[-\int\limits_{[0,x)} r_n(s)\,ds]\,dP(\omega)$.

Claim 1 is a direct consequence of the above definitions. Claim 2 follows since $r_n(s)\xrightarrow{a.s.} r(s)$, $I_{A_n}(\omega)\xrightarrow{a.s.} I_A(\omega)$ and by LDCT the r.h.s. of (7.10) converges to $\int\limits_{\Omega} \exp[-\int\limits_{[0,x)} r(s)ds]I_A(\omega)dP(\omega) = P(r(t)\varepsilon B, X \geq x)$. To prove claim 3 we show below that (i) $f_n(x) \to f(x)$, (ii) $\phi_n(B,x)f_n(x)\to\phi(B,x)f(x)$, observe from (7.13) and (7.15) that $0 \leq \phi_n(B,x)f_n(x) \leq f_n(x)$ and note from the definitions (7.8) and (7.12) that $\int\limits_{[x,\infty)} f_n(x)\,dx \to \int\limits_{[x,\infty)} f(x)dx$. Thus claim 3 follows by a generalization of LDCT (see theorem in Royden [1968], pg. 89).

(i) To show $f_n(x) \to f(x)$, we use (7.4) to write

$$f_n(x) = -\frac{d}{dx}[\Pi(1+\beta_i(x - t_i)^+)^{-\Delta\alpha_i}]$$

$$= \Sigma[\Delta\alpha_i(1+\beta_i(x-t_i)^+)^{-\Delta\alpha_i-1}\beta_i I_{[0,x)}(t_i)\prod_{\substack{j=1\\j\neq i}}^{k(n)}(1+\beta_j(x-t_j)^+)^{-\Delta\alpha_j}]$$

$$= [\Pi(1+\beta_j(x-t_j)^+)^{-\Delta\alpha_j}]\Sigma\,\hat{\beta}_i\,I_{[0,x)}(t_i)\Delta\alpha_i$$

$$\to \exp[-\int_{[0,\infty)}\ln(1+\beta(t)(x-t)^+)d\alpha(t)]\int_{[0,x)}\hat{\beta}(t)d\alpha(t)$$

$$= -\frac{d}{dx}\exp[-\int_{[0,\infty)}\ln(1+\beta(t)(x-t)^+)d\alpha(t)]$$

(7.13) $= f(x)$ by (7.8) and Theorem 3.1.

(ii) To show $\phi_n(B,x)\,f_n(x) \to \phi(B,x)f(x)$, consider

$$(7.14)\quad \phi_n(B,x)f_n(x) = -\frac{d}{dx}\int_\Omega \exp[-\int_{[0,x)}r_n(s)ds]\,I_{A_n}(\omega)dP(\omega).$$

The derivative of the integrand in (7.14) is $-r_n(x)\exp[-\int_{[0,x)}r_n(s)ds]$ which is nonpositive and bounded below by the integrable function $-r_n(t_{k(n)},\omega)$. Thus

$$\phi_n(B,x)f_n(x) = \int_\Omega r_n(x)\exp[-\int_{[0,x)}r_n(s)ds]I_{A_n}(\omega)dP(\omega)$$

$$= \int_{B_n(\underset{\sim}{u},\beta)}\{\Sigma\beta_i u_i I_{[o,x)}(t_i)\}\exp[-\Sigma\beta_i u_i(x-t_i)^+]\Pi g(u_i;\Delta\alpha_i)du_i$$

$$= \sum_{j=1}^{k(n)}I_{[0,x)}(t_j)\int_{B_n(\underset{\sim}{u},\beta)}\beta_j u_j \exp[-\Sigma\beta_i u_i(x-t_i)^+]\Pi g(u_i;\Delta\alpha_i)du_i$$

$$= \sum_{j=1}^{k(n)} \beta_j I_{[0,x]}(t_j) \int_{B_n(\underset{\sim}{u},\beta)} \{u_j^{\Delta\alpha_j} \exp[-u_j(1+\beta_j(x-t_j)^+)]/\Gamma(\Delta\alpha_j)\} du_j$$

$$\prod_{i\neq j} \{u_i^{\Delta\alpha_i - 1} \exp[-u_i(1+\beta_i(x-t_i)^+)]/\Gamma(\Delta\alpha_i)\} du_i$$

$$(7.15) \qquad = [\prod(1+\beta_i(x-t_i)^+)^{-\Delta\alpha_i}]\{\sum_{j=1}^{k(n)} \hat{\beta}_j I_{[0,x]}(t_j)\Delta\alpha_j \int_{B_n(v,\beta)} g(v_j;\Delta\alpha_j+1)dv_j \prod g(v_i;\Delta\alpha_i)dv_i\}$$

where $\beta_i u_i = \hat{\beta}_i v_i$. The term in the brackets converges to

$\exp[-\int_{[0,\infty)} \ln(1 + \beta(s)(x-s)^+)d\alpha(s)]$ and the expression in the braces

can be written as $\int_{[0,x]} \{\sum \hat{\beta}_i F(B;\Gamma(\alpha_n + I_{(t_i,\infty)},\hat{\beta})) I_{(t_{i-1},t_i]}(s)\}d\alpha(s)$.

Since $\alpha(s)$ and $\beta(s)$ are bounded for $s \in [0,x]$ and the integrand

converges to $\hat{\beta}(s) F(B;\Gamma(\alpha + I_{(s,\infty)},\hat{\beta}))$ by Lemma 7.1, application

of LDCT yields

$$(7.16) \quad \phi_n(B,x) f_n(x) \to \exp[-\int_{[0,x]} \ln(1+\beta(s)(x-s)^+)d\alpha(s)]\int_{[0,x]} \hat{\beta}(s)F(B;\Gamma(\alpha+I_{(s,\infty)},\hat{\beta}))d\alpha(s)$$

Thus $\phi_n(B,x) f_n(x) \to \phi(B,x) f(x)$ since the limit in (7.16) equals

$\phi(A,x) f(x)$ by Theorem 3.1, (7.8) and (7.9). This concludes the

proof for $m = 1$. For $m = 2$, a similar proof can be given. The

posterior distribution after the first observation is used as the

prior for the second. Thus $\phi_n(B,x)$ and $f_n(x)$ are similarly

defined except that $r(t,\omega)$ is distributed as a mixture of extend-

ed gamma processes. The detailed computations are more cumbersome.

One can use a generalization of an unsymmetric Fubini theorem given by Cameron and Martin (1941) to interchange the order of certain integrals that are encountered. Using LDCT and the result proved for $m=1$, one arrives at the result for $m = 2$. The proof for arbitrary m follows by mathematical induction.

REFERENCES

[1] Antoniak, C. E. (1974). Mixtures of Dirichlet processes
     with applications to Bayesian nonparametric problems.
     Ann. Statist. 1, 1152-1174.

[2] Cameron, R. H. and Martin, W. T. (1941). An unsymmetric
     Fubini theorem. Bulletin of the American Mathematical
     Society, Ser. 2, 47, 121-125.

[3] Doksum, K. A. (1974). Tailfree and neutral random probabilities
     and their posterior distributions. Ann. Prob., 2, 183-201.

[4] Ferguson, T. S. (1973). A Bayesian analysis of some
     nonparametric problems. Ann. Statist., 1, 209-230.

[5] Ferguson, T. S. and Phadia, E. G. (1979). Ann. Stat.,
     7, 167-184.

[6] Kaplan, E. L. and Meier, Paul (1958). Nonparametric estimation
     from incomplete observations. J. Amer. Statist. Assoc.,
     53, 457-481.

[7] Loéve, M. (1963). Probability Theory. D. Van Nostrand.

[8] Mann, N. R., Schafer, R. E. and Singpurwalla, N. D. (1974).
     Methods for statistical analysis of reliability and
     life data. John Wiley and Sons.

[9] Rosenblatt, Murray (1971). Curve estimates. Ann. Math. Stat.,
     42, 1815-1842.

[10] Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian
     estimation of survival curves from incomplete observations.
     J. Amer. Statist. Assoc., 71, 897-902.